DOCUMENT RESUME

ED 041 829 SP 004 075

AUTHOR Madaus, George F.; Airasian, Peter W.
TITLE Placement, Formative, Diagnostic, and Summative Evaluation of Classroom Learning.
PUB DATE 70
NOTE 24p.; Paper presented at annual meeting, AERA, 1970
EDRS PRICE EDRS Price MF-$0.25 HC-$1.30

ABSTRACT

This paper defines four types of classroom evaluation by comparing the evaluation types across nine dimensions: 1) function, 2) time, 3) characteristics of evidence, 4) evidence gathering techniques, 5) sampling, 6) scoring and reporting, 7) standards, 8) reliability, 9) validity. The four types of evaluation, described by the purpose a teacher has for determining, valuing, describing, or classifying some aspects of student behavior, are 1) placement evaluation used to place students according to prior achievement or personal characteristics, at the most appropriate point in an instructional sequence, in a unique instructional strategy, or with a suitable teacher; 2) formative evaluation used to provide the student and teacher with feedback on the student's progress toward mastery of relatively small units of learning to provide information that will direct subsequent teaching or study; 3) diagnostic evaluation for the identification of students whose learning or classroom behavior is being adversely affected by factors not directly related to instructional practices; 4) summative evaluation used principally to certify, assign a grade, or to attest to the student's successful completion of a relatively large unit of instruction. (Included are charts comparing the four types of evaluation on each of the nine characteristics.) (JS)

Bol. doc.

MAY 13 1970

PLACEMENT, FORMATIVE, DIAGNOSTIC, AND SUMMATIVE EVALUATION OF CLASSROOM LEARNING*

ED041829

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

George F. Madaus
Peter W. Airasian

Boston College

From the mid thirties until the early sixties, primarily as a result of the writings of Ralph Tyler (e.g. 1934, 1950) the emphasis in evaluation was concentrated on the teacher and her unique instructional objectives. Two events were instrumental in shifting the focus in the evaluation literature away from the individual teacher. The first was the advent, during the late fifties and early sixties, of new curriculum development projects, especially in the physical sciences. The appearance of these projects generated concern about the role of evaluation in course development (e.g. Cronbach, 1963; Scriven, 1967; Stake, 1967; Grobman, 1968).

The second event, while harder to pinpoint in time, is no less a reality. It is the growing recognition that the busy teacher responsible for varied work of large and varied classes seldom has the time to carry out individually the operations called for in the Tyler Rationale (e.g., Jackson, 1965; Madaus, 1969).

Despite this shift in the literature, evaluation of some kind is a pervasive and crucial feature of all teaching. Some teacher

---

*Paper presented to the 1970 Annual Meeting of the American Educational Research Association, Minneapolis, Minnesota, March 3, 1970.

SP 001 075

evaluation is spontaneous, unsystematic, and informal, for the most part based upon such cues as momentary facial expressions, shifts in posture, tone of voice, etc. On the other hand, some teacher evaluations are based upon more systematic and quantitative data, derived principally from paper and pencil tests.

The purpose of this paper is to define four types of classroom evaluation (placement, formative, diagnostic, and summative) by comparing these evaluation types across nine dimensions (function, time, characteristics of evidence, evidence gathering techniques, sampling, scoring and reporting, standards, reliability, and validity). The intent of this paper is not to imply that the overburdened teacher should be expected to cope with the requirements of the four types of classroom evaluation. In fact, the very act of outlining and compiling these four types has convinced the authors of the need for cooperative efforts on the part of teachers and school systems if the potential to improve instruction inherent in evaluation is ever to be realized.

The first distinction between the four types of evaluation resides in the purpose a teacher has for determining, valuing, describing or classifying some aspects of student behavior. Figure 1 contrasts the various purposes of placement, formative, diagnostic and summative evaluation.

As the name implies, placement evaluation is used to place students. Based upon his prior achievement or personal characteristics, a student can be placed at the most appropriate point in an instructional sequence, in a unique instructional strategy, or with

- - - - - - - - - - - - - - - - -

Insert Figure 1 here

- - - - - - - - - - - - - - - - -

a suitable teacher. The following analogy is useful to illustrate the concept of placing the student at the optimum point in an instructional sequence. Picture each of the prerequisite skills and anticipated objectives of a course as units on a number line. Course specific or course independent prerequisite skills are analogous to negative numbers, while the presence of these skills but the absence of student mastery of any of the anticipated objectives of the course is analogous to the zero point. The objectives of the course are analogous to the positive numbers along the line. A primary purpose of placement evaluation is to locate a student on this "instructional number line." This analogy limps as these prerequisite skills or the course objectives are not necessarily sequential or hierarchical. However the point is that in many, if not in most schools students are in fact "placed" at our imaginary zero point without regard to their prerequisite skills or prior mastery of course objectives.

Matching a student with an instructional method or with a particular teacher is still in its infancy. However, as research on the efficacy of such placement becomes more abundant, it may be possible to place students either with the most appropriate teacher or in the optimal instructional strategy.

The main purpose of formative valuing is to provide the student and teacher with feedback on the student's progress towards mastery

of relatively small units of learning. Formative evaluation is not used to grade students. Instead, its primary function is to provide information that will direct subsequent teaching and/or study. The function of summative evaluation on the other hand is principally to certify, assign a grade or to attest to the successful completion by the student of a relatively large unit of instruction. Summative information gathered at the end of a relatively large unit of instruction can be used to judge the effectiveness of the teacher's performance in assisting students to realize the course objectives.

The terms relatively large and relatively small are admittedly vague and de facto take their definitions from teacher practice or school policy. For example, formative evaluation could take place daily or weekly; some teachers may give summative exams bi-weekly or monthly. In countries like Ireland and India summative evaluations in the form of Intermediate or Leaving Certificate Examinations take place only after two years of instruction.

The purpose of the evaluation (e.g. to remediate past instruction or to plan future instruction or to grade or certify) rather than the size of the instruction unit is the principal issue.

The function of diagnostic evaluation is the identification of students whose learning or classroom behavior is being adversely affected by factors not directly related to instructional practices. The teacher must be able to recognize factors which are in a sense 'extra-classroom' but nevertheless adversely affect the child's performance in school.

Insert Figure 2 here

The next point of comparison between four types of evaluation lies along a time dimension. Figure 2 contrasts the time points at which evidence is gathered for placement, formative, summative and diagnostic evaluation. Placement evaluation occurs prior to the beginning of a course or an instructional unit. Of course a student may be 'replaced' during the year if the original placement proves, for one reason or other, to be less than ideal. However, this restreaming or regrouping will most likely be the result of formative feedback or summative grades. Formative and diagnostic evaluations take place as instruction unfolds, while summative evaluation because of its grading or certifying, function takes place at the conclusion of an instructional unit.

Unlike other types of evaluation, diagnostic evaluation is a continual act which admits to no exact time constraints. The teacher should always be sensitive to the manifestation of behavioral symptoms assumed to be related to 'extra classroom' causes of learning difficulties.

Insert Figure 3 about here

Figure 3 contrasts the four types of evaluation according to the behavioral characteristics of the evidence gathered. These behavioral characteristics will further differ <u>within</u> a type of evaluation according to the purpose of the evaluation. Across evaluation types, Figure 3 shows that formative, summative and two types of placement evaluation, (namely determining a student's attainment of either prerequisite skills or prior mastery of course objectives) generally collect cognitive or psychomotor data.

Placement evaluation may sometimes seek affective data if its purpose is to match students with certain characteristics with either a certain type of teacher or with a certain mode of instruction. Summative evaluation should gather affective data if the course contains affective objectives. However, individuals should probably not be graded on the basis of such a summative evaluation. The proper objective of affective summative evaluation is to determine the degree to which the class as a whole has attained these objectives. Therefore, anonymously gathered data about the class' attainment of affective objectives is the proper aim of inquiry. Anonymity permits safer inferencesto be made from the data. No reference to affective evidence is made under formative evaluation. This is due solely to the fact that nothing is yet known on either the methodology required by or the consequences resulting from such a practice. However, the guidelines outlined for summative evaluation of affective behavior would likely hold as well for formative evaluation. That is, the data should be gathered anonymously and used to make judgments about group rather than individual progress.

In Figure 3 the behavioral characteristics of the evidence gathered during diagnostic evaluation do not fall under the taxonomic categories of cognition, affect, or psychomotor behavior, but rather are classified as physical, psychological, or environmental in nature. The physical or biological category may include problems of vision, speech, or general health. Psychological symptoms involve emotional or social maladjustment while under the category of environment we find such things as dietary problems or a disrupted or disadvantaged home life.

A final note before leaving Figure 3: Formative or summative evidence should not necessarily be limited to data about course objectives. Evidence should also be obtained about unintended outcomes, both positive and negative, which always accrue during a course.

- - - - - - - - - - -

Insert Figure 4

- - - - - - - - - - -

Figure 4 compares the techniques used during each of the four types of evaluation to gather evidence. Since placement evaluation has a variety of purposes the techniques employed to gather evidence vary. Commercially available intelligence, achievement and diagnostic tests can be used in placing a student. In addition,to standardized tests,locally constructed instruments are generally needed for proper placement. Standardized tests sample objectives that cut across curricula and consequently are often not the most parsimonious means of obtaining information specific enough for local placement needs. Placement data need not result solely from administering paper and pencil instruments. Information relevant to placement decisions may also be obtained by check lists, interviews, observations, etc.

In formative evaluation the predominant technique used to gather evidence is that of locally constructed achievement tests. These tests should be tailored to evaluate student progress through a relatively short unit of instruction. Information gathered from formative achievement tests can, and very often should, be supplemented by interviews, classroom observations, video tapes, teacher intuition, etc. Summative evaluation gathers evidence for grading or certifying

primarily through the use of achievement tests. These tests are most often locally constructed norm referenced tests. Summative tests can be external examinations and in some situations such as in nursing or vocational education can be criterion referenced performance tests.

For diagnostic evaluation many schools routinely employ general screening techniques to identify students with auditory or visual problems. However, the primary technique used to identify students experiencing learning problems resulting from extra-classroom causes is that of sensitive classroom observations by the teacher. Once a teacher observes tell-tale symptoms the correct procedure is generally to refer the student to expert assistance.

- - - - - - - -

Insert Figure 5

- - - - - - - -

Figure 5 compares the four types of evaluation according to the sampling considerations involved in evidence gathering. The sampling considerations in placement evaluation depend on the type of placement sought. The determination of the presence of prerequisite entry behaviors necessitates sampling each prerequisite skill. If the aim is placement in a particular type of instruction or with a particular teacher a sample that ensures a reliable measure of the behaviors associated with the classification construct must be obtained.

Although summative tests are primarily used for grading and certification, they can also be used for placement. If the student obtains a sufficiently high score on a summative pre-test, he may be placed out of the course. If he does not obtain a sufficiently high score, nonetheless the test results may help to determine the optimum starting point in the course. More specific placement information can

be obtained through the use of formative pre-tests.

When formative and summative instruments are used for placement the sampling considerations involved are identical to those for regular formative and summative evaluations. Summative tests are made up of a weighted sample of items designed to measure over-all course objectives; the number of items per objective vary according to the value placed on the particular objective. This valuing may be a function of instructional time, teacher judgments, perceived future value, etc. The point is that summative tests reflect a weighted judgment about the worth of each objective contained in the master table of course specifications.

There are two sampling considerations for building a formative test. The objectives of some formative units build on one another. In such cases each objective in the unit must be sampled in order to determine where in the hierarchy of objectives the student is experiencing difficulty. In other units the objectivesmay be discrete, that is, unrelated, to one another. When this is the case, value judgments similar to those discussed in the preceding paragraph must be made before sampling items.

Since observations are gathered in an ad hoc manner in diagnostic evaluation sampling in the strict psychometric sense is not applicable. It may be that the tell-tale symptoms do not regularly manifest themselves. Further,to wait for further occurrences may retard remedial action. The best approach for a teacher who suspects extra classroom causes to be at the root of learning disorders is to talk to the appropriate referral agency about her observations and

hypotheses. The expert could then either see the child himself or direct the teacher to look for additional behavioral symptoms.

- - - - - - - - - - - - - - - - -

Insert Figure 6 here

- - - - - - - - - - - - - - - - -

6.

Figure 6 distinguishes the scoring and reporting procedures employed by each of the four types of evaluation. In placement evaluation, except when a student places out of a course, results are reported in terms of profiles, patterns or sub-scores on the objectives or characteristics in question. In scoring for placement purposes the unit of analysis which provides the most appropriate data must be carefully chosen. For example, a standardized diagnostic battery may be simply scored as directed; a summative achievement test may be scored in terms of course objectives; a formative test in terms of a student's performance on each test item.

Since the results of formative evaluation are used to direct teachers and students, the fnformation must be highly specific. Consequently scoring and reporting are based on item response patterns. Since students must be free to make mistakes on formative tests without being penalized, scoring and reporting must avoid any indication of ranking or grading.

Diagnostic reports should contain an anecdotal record of the teacher's observations. The concept of a score per se is not applicable. Scores resulting from summative evaluations are typically expressed as the number of items answered correctly. For purposes of

reporting, the raw score is generally converted to letter grades, percentage of correct responses, percentiles, standard scores, stanines, etc.

- - - - - - - - - - - - - - -

Insert Figure 7

- - - - - - - - - - - - - - -

Scores by themselves are often meaningless. A set of standards against which to compare a derived score is also needed. Figure 7 shows that each of the four types of evaluation employs a different set of standards in keeping with differences in function or purpose. The standards employed in placement evaluation are perhaps the most varied. When comparing a student's performance to the performance of previous classes the standard is norm referenced. When determining whether the student has the necessary prerequisite skills the standard can be absolute; that is criterion referenced. When an attempt is made to match students either with a particular teacher or with a particular type of instruction, standards derive either from available research evidence or from the teacher's past experience.

A criterion-referenced standard is used in formative judgments. Formative evaluation compares item response patterns to a predetermined level of mastery for the unit. This level of mastery may be a simple pass-fail criterion or it may be more complex and subjective,based on the teacher's judgment of what constitutes an adequate performance.

Summative evaluation, on the other hand, generally compares a student's score against the performance of a well defined group,

generally the class itself, in an attempt to grade, certify, or select. Since the intent is to differentiate between students, the standards are norm referenced. The standards against which the diagnostic reports are compared are lists or descriptions of behaviors assumed to be related to learning or classroom difficulties.

- - - - - - - - - - - - - - - - -

Insert Figure 8

- - - - - - - - - - - - - - - - -

The reliability of the evidence gathered under each evaluation approach is shown in Figure 8. In placement evaluation, where a broad range of instruments and procedures can be employed, reliability may be the function of the trait being measured, or the consequences of the judgments. In cases in which the intent is to place a student at the proper instructional point, after which there is little latitude to replace the student, the consequence of the placement decision is grave. Thus a very high reliability is required of the instruments used to gather such data. When the placement decisions can be readily modified and systematic regrouping is possible then the reliability considerations can be less stringent.

In formative evaluation, reliability involves the stability or consistency of item response patterns. These response patterns must be demonstrated to be stable and consistent if instructional decisions are to be made with any degree of confidence.

The reliability sought in diagnostic evaluation involves the recurrence of behavioral symptoms. However it should be recognized that observed symptoms can either disappear or become more pronounced over time. Therefore, our use of the term recurrence does not necessarily connote stability or consistency.

Errors in placement or in formatively evaluating students can generally be rectified with relative ease. In diagnostic evaluation there is generally less harm in making an incorrect referral than in failing to refer at all. However, summative decisions are generally final. The results are likely to follow the student throughout his scholastic career. As a consequence, summative scores should be highly reliable, based on achievement tests possessing a high degree of internal consistency and scorer objectivity.

- - - - - - - - - - - - - - - - - -

Insert Figure 9

- - - - - - - - - - - - - - - - - -

The final comparison concerning validity is detailed in Figure 9. Since our four evaluation types deal with classroom instruction, the principal consideration is whether or not the instruments have content validity; that is whether they measure the objectives of instruction.

Less central, yet important, is the construct validity of placement and formative instruments. Matching students either to teachers or to an instructional mode involves a construct or constructs hypothesized to be related to optimum placement. Similarly, the construct validity of a formative instrument which purports to measure a hierarchy of objectives can be tested by determining whether students who fail an item testing a particular objective fail all succeeding items testing dependent objectives.

To discuss validity in diagnostic evaluation we have resurrected the term "face validity." This is not because the term itself is important, but rather because it is one familiar to most evaluators and because it describes in a brief manner the characteristic of the validity involved. The symptoms observed by the teacher are valid if they appear to be symptoms of psychological, physical, or environmental causes of learning disability. Teachers are not trained psychologists, social workers, or nurses. The teacher's prime function is to recognize symptoms. It is the specialist's task to determine whether teachers' observations are in fact valid.

Summary

This paper has defined four types of classroom evaluation by contrasting the types across nine dimensions. A final, Summary Figure brings together all of the comparisons discussed in the paper. Once again, our intent is not to suggest that an individual teacher be responsible for the development and implementation of such a complete evaluation system. Nor is it our intent to suggest that the individual teacher should disregard a formal system of evaluation in favor of the more spontaneous and informal evaluation practices which have been operative for so long. What is needed is a careful consideration of how the four types of evaluation discussed in this paper can be brought within the grasp of the individual teacher.

References

Cronbach, Lee J. "Course improvement through evaluation" Teacher College Record, 1963, 64, 672-683.

Grobman, Hulda. Evaluation activities of curriculum projects." AERA Monograph Series on Curriculum Evaluation, Vol. 2. Chicago: Rand McNally & Co., 1968.

Jackson, Philip W. & Belford, Elizabeth. "Educational 6bjectives and the joys of teaching." The School Review, 1965, Vol. 73, No. 3.

Madaus, George F. "The cooperative development of evaluation systems for individualized institutions" Keynote address to the Annual Southeast Invitational Conference on Measurement in Education, University of Tennessee, Knoxville, Tennessee, October 3, 1969, Boston College: mimeographed.

Scriven, Michael. "The methodology of evaluation" AERA Monggraph Series on Curriculum Evaluation, Vol. 1. Chicago: Rand McNally & Co., 1967, 39-83.

Stake, Robert E. "The countenance of educational evaluation," Teacher College Record, 1967, 68, 523-540.

Tyler, Ralph, Constructing Achievement Tests. Columbus: Ohio State Press, 1934.

Tyler, Ralph. Basic Principles of Curriculum and Instruction: Syllabus for Education 305. Chicago: University of Chicago Press, 1950.

FIGURE 1

FUNCTION OF EVALUATION

Placement	Formative	Diagnostic	Summative
to place students by a. determining the degree to which prerequisite entry behaviors or skills are present or absent b. determining entering mastery of the course objectives c. matching students to alternative teachers or instructional modes according to characteristics known or thought to optimize achievement	Formative evaluation is contributory. Its functions are a. to provide on-going feedback to the teacher for the purposes of 1. choosing or modifying subsequent learning experiences 2. prescribing remediation of group or individual deficiencies b. to provide on-going feedback to the student for the purpose of directing advanced or remedial study	to recognize psychological, physical, or environmental symptoms manifested by students with extraordinary or recurrent learning and/or classroom problems	to grade, certify, or attest to student learning or teacher effectiveness

FIGURE 2

TIME OF EVIDENCE GATHERING

Placement	Formative	Diagnostic	Summative
prior to entry into an instructional unit	several times prior to the completion of instruction on a predefined segment (unit, chapter, etc.) of a course	While a teacher should always be sensitive to the manifestation of symptoms known to be related to learning difficulties, he should be particularly attentive to students when classroom or learning difficulties cannot be explained in terms of cognitive or instructional variables.	at the conclusion of a unit, course, or year's instruction

FIGURE 3

BEHAVIORAL CHARACTERISTICS OF EVIDENCE GATHERED

Placement	Formative	Diagnostic	Summative
dependent on the functions stated in Figure 1: typically cognitive or psychomotor as is generally the case when the function is to determine whether or not prerequisite entry behaviors are present or to determine the student's entry mastery of course objectives, may also be affective as is sometimes the case when the purpose is to match students to alternative teachers or instructional modes.	cognitive or psychomotor	physical (vision, auditory perception, dominance and laterality, general health, etc.); psychological (intelligence, emotional maladjustment, social maladjustment, etc.); environmental (nutritional, parent-child relationships, peer influences, etc.)	depends on course objectives: higher or lower level cognitive behaviors, affective, and/or psychomotor

FIGURE 4

EVIDENCE GATHERING TECHNIQUES

Placement	Formative	Diagnostic	Summative
depends on type of placement sought, but could include: a. commercial tests (intelligence, achievement, diagnostic, etc.) b. teacher-made instruments (formative, summative, specially designed pre-tests, observation, interviews, checklists, video-tapes, etc.)	a series of teacher-made achievement measures: supply, essay, or selection tests; interviews, video-tapes, checklists, etc.	primarily observational although for certain symptoms general screening techniques to confirm hypotheses may be available to the classroom teacher (eg. vision). Generally, upon noting symptons, the teacher forwards his observations to proper agencies, eg., guidance counselor, nurse, school psychologist, etc.	primarily internally or externally constructed achievement tests

FIGURE 5

SAMPLING CONSIDERATIONS FOR EVIDENCE GATHERING

Placement	Formative	Diagnostic	Summative
depending on the function specified in Figure 1, evidence must be gathered on a. each prerequisite entry behavior b. a representative sample of course objectives c. those behaviors related to a construct(s) which in turn are known or thought to be related to different types of teachers or to alternative modes of instruction	a. where the objectives of an instructional segment are interrelated (cognitively or sequentially within cognitive levels) the sample should include all objectives in the segment. b. where the objectives of a predefined instructional segment are not interrelated, the determination to include all or only some of the objectives may depend on such considerations as: the use of an objective in subsequent learning; the extent to which an objective integrates or reinforces prior learning, etc.	Sampling in the psychometric sense is not applicable. an ad hoc observational process designed to construct and confirm hypotheses about suspected causes of disorders	weighted sample of course objectives: weighting may be in terms of teaching emphasis, purpose of evaluation, time spent, transferability of objectives, perceived importance in society, etc.

FIGURE 6

SCORING AND REPORTING

Placement	Formative	Diagnostic	Summative
patterns, profiles, sub-scores, etc. for all functions in Figure 1 except placing students out of a course where total score may be utilized	patterns of item responses, abilities mastered or not mastered, etc. All reporting must be free of any intimation of a mark, grade, or certification.	an anecdotal report containing specific behavioral instances forwarded to the appropriate referral agency	sum of total number of correct responses, either by objective or on total exam; reported as letter or number grade, standard score, percentile, etc.

FIGURE 7

STANDARDS AGAINST WHICH SCORES ARE COMPARED

Placement	Formative	Diagnostic	Summative
predetermined norm or criterion referenced standards	criterion referenced standards	compare manifested behavior against specified abnormal behaviors	almost exclusively norm referenced

FIGURE 8

RELIABILITY

Placement	Formative	Diagnostic	Summative
dependent on the characteristics of the instrument used, the trait measured, the scoring procedure, and the consequences of the judgments	stability and/or consistency of item response patterns	recurrence of behavioral symptoms	internal consistency

FIGURE 9

VALIDITY

Placement	Formative	Diagnostic	Summative
primarily content validity but also construct validity (where students are matched with teachers or instructional strategies)	primarily content validity but also construct validity (where hierarchies of objectives are involved)	face validity	content validity